Computers in Biology and Medicine 184 (2025) 109344

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

SARS-CoV-2 main protease (M-pro) mutational profiling: An insight into mutation coldspots

Pol Garcia-Segura, Ariadna Llop-Peiró, Nil Novau-Ferré, Júlia Mestres-Truyol, Bryan Saldivar-Espinoza, Gerard Pujadas, Santiago Garcia-Vallvé *

Universitat Rovira i Virgili, Departament de Bioquímica i Biotecnologia, Research group in Cheminformatics & Nutrition, Campus de Sescelades, 43007, Tarragona, Spain

ARTICLE INFO

Keywords:
COVID-19
3CL-Pro
M-Pro
Genomic profiling
Mutation coldspots
Dimerization inhibition

ABSTRACT

SARS-CoV-2 and the COVID-19 pandemic have marked a milestone in the history of scientific research worldwide. To ensure that treatments are successful in the mid-long term, it is crucial to characterize SARS-CoV-2 mutations, as they might lead to viral resistance. Data from >5,700,000 SARS-CoV-2 genomes available at GISAID was used to report SARS-CoV-2 mutations. Given the pivotal role of its main protease (M-pro) in virus replication, a detailed analysis of SARS-CoV-2 M-pro mutations was conducted, with particular attention to mutation-resistant residues or mutation coldspots, defined as those residues that have mutated in five or fewer genomes. 32 mutation coldspots were identified, most of which mediate interprotomer interactions or funneling interaction networks from the substrate-binding site towards the dimerization surface and *vice versa*. Besides, mutation coldspots were virtually conserved in all main proteases from other CoVs. Our results provide valuable information about key residues to M-pro structure that could be useful in rational target-directed drug design and establish a solid groundwork based on mutation analyses for the inhibition of M-pro dimerization, with a potential applicability to future coronavirus outbreaks.

1. Introduction

Coronaviruses (CoVs) are a large group of enveloped positive-sense ssRNA viruses. Phylogenetically, they belong to the order *Nidovirales* and family *Coronaviridae*. The group is made up of four genera: namely, *Alphacoronavirus* (αCoV), *Betacoronavirus* (βCoV), *Gammacoronavirus* (γCoV) and *Deltacoronavirus* (δCoV) [1]. Specifically, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is clustered within βCoVs. Prior to the outbreak of SARS-CoV-2, only six CoVs –two αCoV (*i.e.,* hCoV-229E and HKU-NL63) and four βCoVs (*i.e.,* hCoV-OC43, hCoV-HKU1, SARS-CoV and MERS-CoV)– could infect humans [2]. Of these, only SARS-CoV and MERS-CoV caused severe lower respiratory tract infections and extrapulmonary manifestations, which eventually resulted in death, similar to those reported in COVID-19 [3,4]. Phylogenetic analyses have shown that SARS-CoV-2 is found in a sister clade to the SARS-CoV and bat SARS-related-CoVs (SARSr-CoV) and the closest known neighbor is the bat coronavirus RaTG13 with an overall sequence identity of 96.2 % [5,6]. Interestingly, although closely related, SARS-CoV-2 is clearly distinct to SARS-CoV and MERS-CoV with a genomic similarity of about 80 % and 50 %, respectively [7]. A more detailed analysis at protein level shows that generally SARS-CoV-2 is very similar to SARS-CoV, especially in terms of the important proteins such as the main protease (96 % identity), the envelope protein (95 % identity) or the nucleoprotein (94 % identity). Nonetheless, other proteins such as the spike protein or the papain-like protease (PL-pro) are less similar to their homologues in SARS-CoV with 76 % identity in both cases [1,8].

SARS-CoV-2 has a very compact genome of about 29.9 kb in size, encoding >9000 residues. The genetic makeup of SARS-CoV-2 contains two flanking untranslated regions (UTRs) consisting of a 5′ cap structure and a poly(A) 3′ end, and 12 genes encoding 25 different proteins in between. Other ORFs can be found within structural protein genes [1,9]. SARS-CoV-2 –and other CoVs, in general– present two overlapping ORFs (*i.e.,* ORF1a and ORF1b) encoding for two polyproteins, pp1a and pp1ab, respectively. Polyproteins require an autoproteolytic cleavage to give rise to the 16 non-structural proteins that form the replicase/-transcriptase complex. Two proteases within this non-structural protein –namely, the main protease (M-pro) and the PL-pro– mediate this vital

* Corresponding author.
E-mail addresses: polgarse2@gmail.com (P. Garcia-Segura), nnovauf@gmail.com (A. Llop-Peiró), juliamt110@gmail.com (N. Novau-Ferré), ariadna.llop@urv.cat (J. Mestres-Truyol), bsaldivar.emc2@gmail.com (B. Saldivar-Espinoza), santi.garcia-vallve@urv.cat (S. Garcia-Vallvé).

https://doi.org/10.1016/j.compbiomed.2024.109344
Received 9 July 2024; Received in revised form 20 September 2024; Accepted 30 October 2024
Available online 12 November 2024

function. In fact, the M-pro is known to cleave no fewer than 11 peptidyl bonds on the large polyprotein 1 ab including its own N-terminal and C-terminal processing sites [10,11]. The SARS-CoV-2 M-pro, also known as 3-chymotrypsin-like (3C-like) protease, is a 306-residue long protease that is crucial for virus replication. M-pro is being intensively studied as a pharmacological target against COVID-19 [12–15]. Moreover, no homologous proteins in humans are known, which makes the inhibition of this target even more interesting [11]. Three domains can be identified within the M-pro protein: domain I (residues 8–101), domain II (residues 102–184) and domain III (residues 201–303) [10,16]. The substrate-binding site is located at a cleft between domains I and II. Domain III has a globular structure composed of five helices that regulate dimerization, as M-pro is known to be catalytically inactive as a monomer [11]. Thus, domain III is not directly involved in M-pro catalytic activity, but it appears to be indispensable for M-pro function. Dimerization is mediated by interactions between domain II of one monomer and the N-terminal region (N-finger, residues 1–7) of the other chain in a contact interface of ~1394 $Å^2$ [17]. The N-finger must squeeze in between domains II and III of the monomer and domain III of the opposing chain to reach an important residue, Glu166, which is buried in the domain II/III interface. By doing this, the N-finger helps to shape the substrate-binding site [11]. Four important subsites can be identified within the substrate-binding site: S1, S1', S2 and S3. M-pro has a catalytic dyad made up of His41 and Cys145 located deep within the substrate-binding site [16]. Cys145 is part of the oxyanion loop, an S-shaped loop of domain II formed by residues from Gly138 to Gly146. A water molecule, known as catalytic water (H_2O_{cat}), appears to play an important role within the substrate-binding site, probably acting as the missing residue of a canonical substrate-binding site typically found in chymotrypsin-like proteases [16]. For instance, H_2O_{cat} enables the His41 to be correctly positioned through a complex H-bond network also involving Asp187 and His164 [16,17].

The biological process of mutagenesis is a driving force for evolution. Genetic diversity in a virus allows evolutionary pressures to select for mutations responsible for greater viral fitness [18]. Selection can also drive the emergence of mutations that allow the virus to escape vaccines or antiviral drugs. Drug-resistant mutants may arise both in the presence and absence of drug selection pressure [19,20]. Thus, SARS-CoV-2 has co-evolved with its host and new variants have arisen [2,21]. Genomic surveillance allows emerging variants to be monitored and presents the basis for structural and functional characterization of mutations. Viral mutation rates vary widely, especially due to the differences in the fidelity of the polymerases used in replication [22]. RNA viruses which use RNA-dependent RNA-polymerases (RdRp) usually have higher mutation rates, because RdRps are more prone to errors than RNA-dependent DNA-polymerases [22]. Nonetheless, it is interesting how CoVs have a unique feature: mutation rates are significantly lower than those reported for other RNA viruses, presumably due to the 3′-to-5′ proofreading activity provided by nsp14 [23]. This may explain the abnormally large genome observed in CoVs (from 27 to 32 kb), compared to usual ssRNA(+) viral genomes. From a simplistic point of view, it is generally assumed that larger genomes have higher mutation rates. Thus, the proofreading activity of nsp14 might strike a balance between mutations, which indeed favors viral fitness, and the correction of some mutations, thus preventing viable virions from being outnumbered by unviable virions which have acquired too many mutations [23,24]. Viral mutations can also occur because of nucleic acid editing by endogenous deaminases [18]. In mammalian species, the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of deaminases target ssDNA and ssRNA to deaminate cytosines into uracils (C > U) [25], while the adenosine deaminases acting on RNA (ADAR) family acts on dsRNA to deaminate adenines into inosines (A > I) [23,26]. Although some work has been done to monitor the evolution of SARS-CoV-2, especially focusing on the residues that have mutated the most, known as hotspots [21,23,27–39], little is known about the biological reasons why some genomic regions or residues are conserved. Here, a mutational profile of SARS-CoV-2 was used to identify 32 mutation coldspots in the M-pro of SARS-CoV-2. Coldspots, defined as unmutated or less mutated residues, might correspond to important residues of the protein in terms of structure and/or function. The structural implications for most of the above mentioned coldspots, as well as their conservation in the main proteases of other CoVs, are discussed.

2. Materials and methods

2.1. Sequence retrieval

A total of 16,572,590 complete full-length SARS-CoV-2 genomic sequences were downloaded from the GISAID database [40] on February 26, 2024. The quality of the sequences was guaranteed by applying the following filters: (i) only sequences obtained from human samples, with less than 150 mutations were considered (ii) partial sequences were avoided by only considering sequences with a minimum length of 29, 000 bp, and (iii) only sequences labelled as "high coverage" were considered (*i.e.,* sequences containing: (a) less than 1 % of unidentified bases (Ns), (b) less than 0.05 % of unique amino acid mutations, to withdraw possible sequencing artefacts, and (c) no insertions and/or deletions, unless verified by the submitter). For the 5,784,122 sequences that passed the defined filters, mutations were identified relative to a reference sequence using a multiple sequence alignment of all available sequences in GISAID. The complete genome NC_045512.2, isolated from Wuhan and submitted to the GenBank database on January 17, 2020, was used as the reference. It is worth bearing in mind that there is a bias in the genomes being analyzed, as the sequencing rates in different countries vary greatly. For instance, more than half of the genomes deposited were from the UK and the USA. Nonetheless, under no circumstances does this bias invalidate the results reported herein.

2.2. Mutation analysis and M-pro characterization

Mutation frequencies were calculated as the number of a specific mutation in the total number of genomes. Mutations were classified as synonymous mutations (*i.e.,* mutations that do not affect the encoded residue), missense mutations (*i.e.,* mutations that cause an amino acid substitution) and nonsense mutations (*i.e.,* mutations that change the codon to a stop codon). Multiple nucleotide substitutions occurring concurrently in the same codon were considered as simultaneous mutations. All the SARS-CoV-2 mutations analyzed can be found at the SARS-CoV-2 Mutation Portal (http://sarscov2-mutation-portal.urv.cat/) [21]. Mutation coldspots are defined as those residues that have mutated in five or fewer genomes (considering only missense mutations. Synonymous and nonsense mutations were not taken into account for this definition). The value five is the first decile threshold of the distribution of the number of times each residue in the M-pro protein has changed.

The average between both subunits of the difference between the free energy of Ala-mutants (ΔG_{mutant}) and that of the WT (ΔG_{WT}) ($\Delta\Delta G = \Delta G_{mutant} - \Delta G_{WT}$), where ΔG is the Gibbs free energy of folding, were estimated with FoldX v3.0 [41]. The ligand-free dimer structure corresponding to the re-refined and rebuilt structure of the SARS-CoV-2 M-pro (PDB ID: 7ALI) from the PDB-REDO database [42] was used for these calculations. All individual Ala-mutants of the amino acid positions defined here as coldspots were generated. To estimate $\Delta\Delta G$ values, the difference between ΔG of each mutant and that of the wild type was calculated, averaging the results between the two subunits. We obtained therefore a measure of the contribution to protein stability for the M-pro amino acid positions defined as coldspots. Positive $\Delta\Delta G$ values indicate a less stable state when these amino acids are mutated to an alanine.

2.3. SARS-CoV-2 M-pro multiple alignment with other CoVs

To retrieve the sequence and structure of the main protease of

different CoVs used in the multiple sequence alignment, structures with at least 90 % sequence similarity with SARS-CoV-2 M-pro were obtained from the Protein Data Bank. Structures corresponding to SARS-CoV-2 were not considered, apart from the PDBid 6WQF, which was used as the SARS-CoV-2 structure and sequence reference. Then, sequences were aligned with the R package msa v1.34.0 [43] using ClustalW with the default settings as the alignment algorithm. The secondary structure assigned by STRIDE [44] to the PDBid 6WQF was also added.

3. Results and discussion

3.1. SARS-CoV-2 M-pro mutations

From the analysis of 5,784,122 SARS-CoV-2 genomes available as of February 26, 2024, a total of 2,773,472 mutations were identified within the M-pro gene. These mutations encompassed 1,605,141 synonymous mutations, 1,168,176 missense mutations and 155 nonsense mutations. On average, this corresponds to approximately 0.48 mutations per genome. Remarkably, 88.9 % of all analyzed genomes exhibited either no mutations or only one mutation within the M-pro gene (Table S1). At the beginning of the pandemic, the majority of SARS-CoV-2 genomes lacked any mutations in the M-pro gene. However, as expected, the number of mutations within this gene increased over the course of the pandemic, particularly after January 2022, when the omicron variant emerged (Fig. S1). Many of the M-pro mutations were repeatedly observed in multiple SARS-CoV-2 genomes. Consequently, the total count of distinct mutations found in the M-pro gene was 2297, comprising 1667 missense mutations, 576 synonymous mutations and 54 nonsense mutations (Table S2). Single nucleotide substitutions (SNS) were considerably more frequent than double or triple mutations within a codon (Table S2). Additionally, the third position of the codon exhibited the highest frequency of mutations, followed by the first and the second positions (Table S2). Regarding the SNS types, U > C transitions were the most prevalent, accounting for 13.6 % of all SNS, but transversions outnumbered transitions (59.3 % vs 40.7 %, respectively) (Table S2). By dividing the total number of M-pro distinct mutations by the length of the gene, we calculated a mutation rate of approximately 2.5 substitutions per nucleotide. This means that, on average, 2.5 distinct substitutions have been identified for each nucleotide of the M-pro gene. However, compared to other SARS-CoV-2 regions, the M-pro gene has a lower mutation rate [21,32,45,46]. Residue substitutions were widely distributed across the different domains of the M-pro protein [47]. The substrate-binding site had a greater number of unique substitutions per residue compared to other domains of the protein (Fig. 1), indicating that this domain has a certain tolerance to mutations. Notably, one of the most conserved region within M-pro was domain II (Fig. 1). This region is one of the most important domains in M-pro dimerization and mediates crucial interactions of the dimer interface.

The most frequent M-pro mutations were the missense substitution C10449A (P132H at protein level) and the synonymous substitutions G10447A and C10198U (Fig. 2). But none of the three substitutions had a frequency greater than 10 %. The P132H substitution is prevalent in all Omicron subvariants and has been proposed not to alter the enzymatic activity or inhibitor binding. However, it may increase protein flexibility by reducing thermal stability and could expand substrate specificity or alter peptide binding [34]. The next most frequent mutations (with a

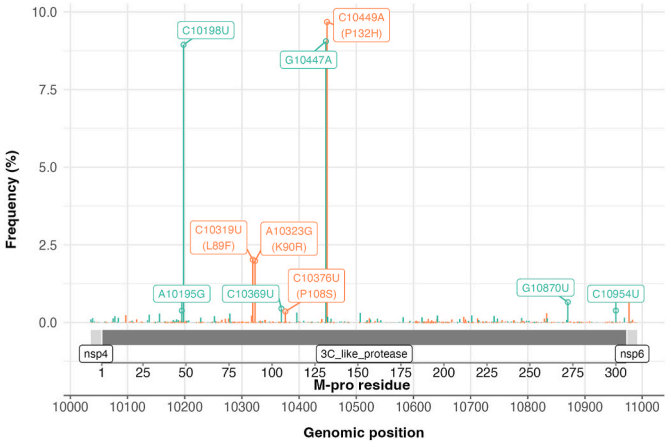

Fig. 2. Mutational profile of the SARS-CoV-2 M-pro gene. The 2297 nucleotide substitutions in the M-pro gene are displayed in their specific genomic position (x axis) and the frequency with which they were found (y axis). Synonymous and missense mutations are colored in turquoise and coral, respectively. The top ten most frequent mutations are labelled with information about the mutation and the residue change in brackets (if applicable).

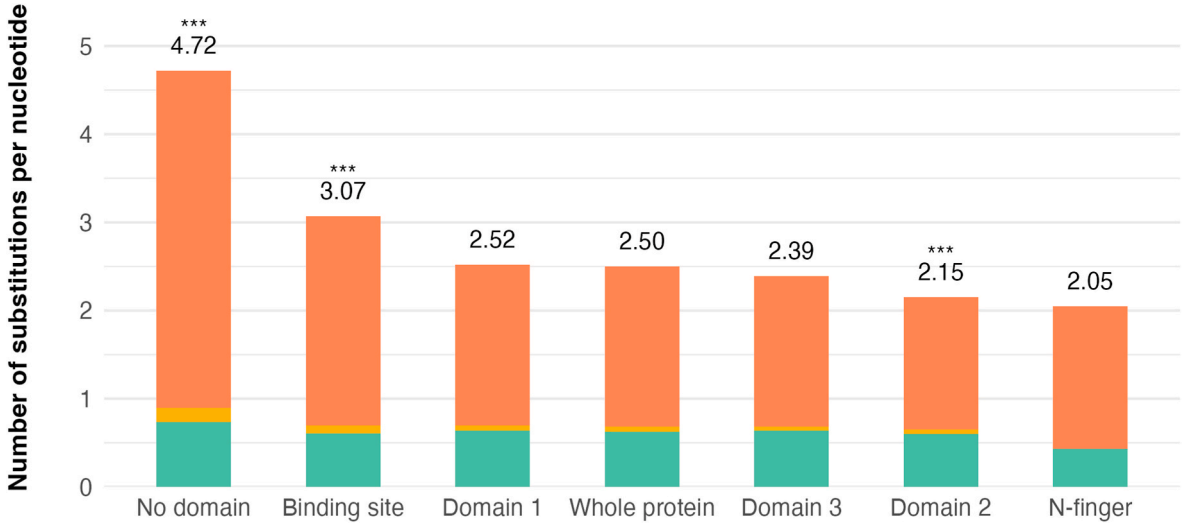

Fig. 1. Number of substitutions found per nucleotide in different domains/regions of the SARS-CoV-2 M-pro gene. Synonymous, nonsense and missense mutations are colored in turquoise, yellow and coral, respectively. The sum of unique mutations per nucleotide in each domain is displayed on top of each bar. The statistical significance when comparing the number of total unique mutations of each domain to the whole protein on top of each bar (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$). Residue spanning of each domain/region: N-finger, 1–7; Domain 1, 8–101; Domain 2, 102–184; Domain 3, 201–303; No domain, 185–200, 304–306; Binding site, 24–26, 41, 46–51, 138–146, 163, 165–168, 172, 186–192.

frequency of about 2 %) in the M-pro gene were the missense mutations A10323G (K90R at protein level) and C10319U (L89F at protein level) (Fig. 2). Mutation A10323G, which results in a Lys to Arg change in residue 90, appeared early during the pandemic, in January 2020, and defines a major clade in SARS-CoV-2 phylogeny according to analyses carried out by the open-source project NextStrain [48]. In fact, this mutation is a defining mutation of the well-known beta variant, designated 20H or B.1.351 [49]. On a protein level, this substitution has been reported to increase the structural stability of the polyprotein pp1ab [50]. The C10319U mutation, which changes Leu89 to Phe, also arose early in the pandemic (January 2020) and it is a recurrent mutation, i.e., that appears multiple times independently and is present in several SARS-CoV-2 lineages [31]. The activity of the L89F mutant has been reported to be comparable to the wild-type protein [51]. Given that Leu89 is a buried residue of domain I, changing it to a bulkier hydrophobic residue, such as Phe, may allow for stronger hydrophobic interactions with surrounding residues (*e.g.,* Val20, Cys38, Leu87) and increase domain I stability. None of the most prevalent M-pro missense substitutions found in this manuscript, such as P132H, K90R and L89F, change significantly the enzymatic activity of M-pro and the efficacy of the specific M-pro inhibitor nirmaltrelvir is not compromised [51].

Table 1
SARS-CoV-2 M-pro mutation coldspots.

Residue	Missense mutations	Nucleotide mutations[a]	Times found	Average FRET Activity[b]	FoldX Average ΔΔG[c] Kcal/mol	Putative function of the residue
Gly2	G2C G2A	G10058U G10059C	1 1	0.85	5.24	Participates in dimer structure
Asn28	N28I	A10137U	2	–	0.1	Essential for enzymatic activity
Phe31	W31R W31L W31F W31C	U10145C G10146U G10146U_G10147U G10147U	2 1 1 1	0.20	4.92	Unknown
Cys44	C44S C44L C44F	U10184A U10184C_G10185U G10185U	1 1 1	0.74	0.53	Forms a SONOS bridge, important for structural stability under oxidative stress conditions
Tyr54	–	–	–	0.42	4.30	Forms the S2 subsite. Essential for enzymatic activity. Participates in substrate binding
Asn63	N63D N63T N63S	A10241G A10242C A10242G	1 1 1	0.99	−0.30	Unknown
Phe66	F66V F66L F66L	U10250G U10250C C10252A	1 1 3	0.46	5.69	Unknown
Gly79	G79R G79T G79E G79V	G10289A G10289A_G10290C G10290A G10290U	1 1 2 1	0.94	0.11	Unknown
Leu115	L115V	U10397G	3	0.07	4.58	Essential for enzymatic activity
Tyr118	Y118H Y118C	U10406C A10407G	1 3	0.28	2.91	Important for enzymatic activity
Tyr126	Y126H Y126D Y126C Y126S	U10430C U10430G A10431G A10431C	1 1 2 1	0.64	6	Form the substrate-binding site
Ser139	S139A	U10469G	3	0.85	−0.32	Participates in dimer structure
Phe140	F140L F140L	U10472C C10474A	2 2	0.31	4.79	Forms the S1 subsite and participates in dimer structure and substrate binding
Gly146	G146S G146C G146V	G10490A G10490U G10491U	1 1 2	0.003	8.99	Essential for enzymatic activity. Impotant for protein structure stability
Ser147	S147G S147I S147R S147R	A10493G G10494U U10495G U10495A	2 1 1 1	0.05	0.05	Essential for enzymatic activity. Important in dimer structure formation
Gly149	G149S G149C G149V	G10499A G10499U G10500U	1 1 1	0.09	2.81	Essential for enzymatic activity
Phe150	F150Y F150C F150L	U10503A U10503G U10504A	1 1 1	0.1	4.63	Essential for enzymatic activity. Important to form the binding site
Tyr161	Y161D Y161H Y161N Y161C	U10535G U10535C U10535A A10536G	1 1 1 1	0.33	4.77	Interacts with His163
His163	H163L	A10542U	1	0.01	1.60	Essential for enzymatic activity. Participates in substrate binding
His164	H164N H164P	C10544A A10545C	3 1	0.63	2.11	Forms the S1 subsite, interacts with H_2O_{cat} and participates in substrate binding
Leu167	L167I L167F	U10553A A10555C	1 2	0.21	2.51	Forms the S4 subsite and participates in substrate binding
His172	H172Y	C10568U	5	0.26	2.28	Forms the S1 subsite and participates in dimer structure and substrate binding
Gly174	G174V	G10575U	1	0.002	8.11	Essential for enzymatic activity. Impotant for protein structure stability
Asp176	D176N D176V D176E	G10580A A10581U_C10582U C10582A	2 1 1	0.33	4.29	Unknown
Gly179	G179S G179C G179A	G10589A G10589U G10590C	2 2 1	−0.004	9.99	Essential for enzymatic activity. Important for protein structure stability
Tyr182	–	–	–	0.23	5.07	Important for enzymatic activity
Trp207	W207R W207L W207C W207C	U10673C G10674U G10675U	1 1 1 2	0.51	3.36	Participates in dimer structure
Asn231	N231D N231G N231T N231I	A10745G A10745G_A10746G_C10747U A10746C A10746U	1 1 1 1	0.94	0.75	Unknown
Glu290	E290K E290D	G10922A A10924U	1 1	0.005	2.93	Participates in dimer structure
Phe291	F291Y F291L	U10926A U10927G	2 1	0.78	2.56	Participates in dimer structure
Gln299	Q299T Q299L	C10949A_A10950C A10950U	1 2	0.38	1.37	Participates in dimer structure
Gln306	Q306R	A10971G	1	1.00	0	Formation of an M-pro cleavage site.

[a] Mutations found together in the same codon are separated by _.
[b] Average FRET normalized scores from the mutational scan of the M-pro that analyzed the function of all possible single amino acid changes. Scores are normalized with the wild-type (WT) M-pro activity set to 1 and the activity of the M-pro containing a stop codon in this position set to 0 (data from Ref. [39]).
[c] ΔΔG are the average between both subunits of the difference between the free energy of Ala-mutants (ΔG_{mutant}) and that of the WT (ΔG_{WT}) ($\Delta\Delta G = \Delta G_{mutant} - \Delta G_{WT}$), calculated with FoldX.

3.2. M-pro mutation coldspots

Mutation coldspots (*i.e.*, residues ranked in the first decile with less missense mutations annotated) may be important, at a structural and functional level, residues for a protein. Therefore, identifying them could be useful for defining putative binding sites for antivirals. To this end, given the key role of M-pro in SARS-CoV-2 replication, it might be interesting to define mutation coldspots in this target. Table 1 lists the 32 M-pro coldspots found in the present study. The residues that do not match the criteria to be categorised as coldspots, but have missense mutations found in total between six and eleven times are here considered conserved residues (Table S3). We expected that the two catalytic residues (*i.e.,* His41 and Cys145) were among the most conserved residues regarding missense mutations. Mutations in theses two residues have been documented, but at very low frequency. We found Cys145 among the conserved residues, with four missense mutations present in a total of seven genomes (Table S3 and Table S4). For His41, we found six missense mutations in a total of 156 genomes (Table S5). However, 147 of them were H41Q mutations found between April and May 2022 in the same region and submitted by the same laboratory (Table S5). His41 and Cys145 mutations are highly unlikely to be present in clinical isolates [47], because altering these residues inactivates the M-pro and would completely compromise viral replication [52–55]. Missense mutations in His41 and Cys145 could be considered therefore a sequencing artefact. In fact, as reported by Jacot et al. [56] few laboratories reported quality criteria of sequences before they were submitted to public databases. Therefore, it is likely that low-frequency variants might be the result of sequencing artefacts, cross contamination between samples or the presence of distinct SARS-CoV-2 genomes in the sample. However, it has been described that at least the mutant C145S generates an active but much slower version of M-pro [57].

Table 1 also shows the average FRET normalized scores derived from the mutational scan of the M-pro, which analyzed the impact of all possible single amino acid substitutions into the enzymatic activity of the protein [34]. FRET scores were normalized with the wild-type (WT) M-pro activity set to 1, and the activity of the M-pro containing a stop codon in this position was set to 0 [34]. Mutations at coldspots and conserved residues generally exhibit lower average FRET scores than the rest of M-pro residues (Fig. S2 and Fig. S3). In fact, 23 out of the 32 coldspots have a FRET normalized score lower than 0.7 (Table 1), meaning that the average of all possible mutations in these positions reduce the enzymatic activity of the SARS-CoV-2 M-pro. Notably, the coldspots residues Leu115, Gly146, Ser147, Gly149, Phe150, His163, Gly174, Gly179, and Glu290 are indispensable for M-pro activity, as any mutation of these amino acids drastically reduces the M-pro enzymatic activity (Table 1) [34]. While functional importance usually predicts accurately evolutionary conservation, conservation does not always predict functional significance. Some residues are strongly conserved but can be mutated without a significant impact on function [34]. An example of this is the conservation of the last residue of the M-pro, the coldspot Gln306, which is part of the M-pro cleavage site itself. Mutations in this amino acid do not significantly affect the protein's activity (Table 1), but they may prevent the ability of M-pro to self-cleave.

Table 1 also shows the average ΔΔG values, i.e., the difference between the Gibbs free energy of folding of Ala-mutants (ΔG_{mutant}) and that of the WT (ΔG_{WT}), calculated with FoldX. These values represent the contribution of residues defined as coldspots to the protein structure stability. Positive ΔΔG values imply a less stable state when these amino acid are mutated to alanine. Most coldspots have positive ΔΔG values (Table 1), demonstrating that a mutation of these residues to alanine destabilizes the protein structure. Notably, the contributions of Gly146, Gly174, and Gly179 are significant, with ΔΔG values exceeding 8 kcal/mol. Although the results regarding amino acid conservation and the effect of their mutation on protein stability are generally consistent, Gln306 does not influence protein stability (Table 1). Its conservation is attributed to the formation of an M-pro cleavage site.

Among the M-pro coldspots there are some of the residues, such as Tyr54, Phe140, His163, His164, Leu167 and His172 (Table 1), that directly bind one of M-pro's natural substrates or inhibitors [13,53,54]. Mutations in most of these residues confer resistance to nirmatrelvir, a specific antiviral drug that target the SARS-CoV-2 M-pro [35,39,58]. However, other substrate-binding or inhibitor-binding residues, such as Thr25, Thr26, Leu27, Met49, Leu141, Asn142, Gly143, Ser144, Met165, Glu166, Pro168, Asp187, Arg188, Gln189, Thr190 and Gln192, are not found among the M-pro coldspots, showing that the M-pro substrate binding domain has a certain tolerance to mutations. Surprisingly, the majority of the coldspots are not located at the substrate-binding site (Fig. 3A), but rather they mediate interactions related to dimerization of the main protease.

To further investigate the biological importance of M-pro mutation coldspots, we decided to assess the conservation of M-pro in different CoVs (Fig. 4). To this end, a multiple sequence alignment was performed. M-pro is a highly conserved protein among various CoVs, with conserved clusters, such as the N-terminal domain and domain II (which contains 18 out of the 32 mutation coldspots), some residues belonging to the substrate-binding site and the catalytic His41 and Cys145 (Fig. 4). The most variable regions include the second part of domain I and the first part of domain III. Virtually all mutation coldspots (28 out of 32) were conserved in at least 50 % of the sequences considered and 22 out of 32 were totally conserved in all CoVs analyzed (Fig. 4), highlighting again the importance of these residues.

More than four years of mutations might be enough for the virus to accumulate some key mutations, but also to maintain residues which are important for its replication machinery. Krishnamoorty and Fakhro [59] published a study of M-pro residues with no reported mutations. However, their work had some limitations: the data was downloaded in early November 2020, and it contained only 19,154 genomes with mutations in the M-pro. Here, given the increase in the sequencing rate in mid-November 2020 we decided to further investigate the role of these presumably crucial M-pro residues. It is worth mentioning that some of the knowledge presented comes from previous studies using the SARS-CoV M-pro but, given the high identity and structural similarity between it and the SARS-CoV-2 M-pro, many of the results may be extrapolated. Fig. 3A and Table 1 show the 32 mutation coldspots identified in the SARS-CoV-2 M-pro in the present study. Only residues mediating known key interactions and/or those for which experimental results have been reported will be discussed below.

3.2.1. Residues from the N-finger and their interaction networks

The first amino acid of the M-pro, a serine, is part of its own cleavage site, defined as a glutamine at the P1 position and a small amino acid (such as Ser, Ala, or Gly) at the P1' position. However, Ser1 is not classified as a mutation coldspot due to 305 observed missense mutations transforming serine into glutamine. Nonetheless, the presence of glycine at position 2 probably allows the resultant QG sequence to be recognized and cleaved by Mpro, serving as a putative functional cleavage site. In addition to its role as a cleavage site, Ser1 also holds structural importance. Ser1 from one protomer is involved in shaping the substrate-binding site of the other protomer, primarily via a well-known salt bridge between its N-terminal amino group and the side chain of Glu166, and an intermolecular H-bond with the backbone carbonyl of Phe140, both from the other protomer (Fig. 3B) [12,60]. Here, Phe140 is identified as a mutation coldspot (Table 1), while Glu166 is a conserved residue because it has three different missense mutations found in seven cases in total (Table S3). This agrees with mutational scans that have been found that Phe140 is intolerant to mutations [54] and Glu166 is tolerant to some of them [53,54] (Fig. S2). However, the role of Glu166 has been reported to be crucial because it mediates an important connection of the substrate-binding site with the dimer interface [60,61]. In fact, mutations of Glu166, in an R298A mutant, block substrate-induced dimerization of M-pro *in vitro* and results in a complete loss of enzymatic activity [61]. In the majority of

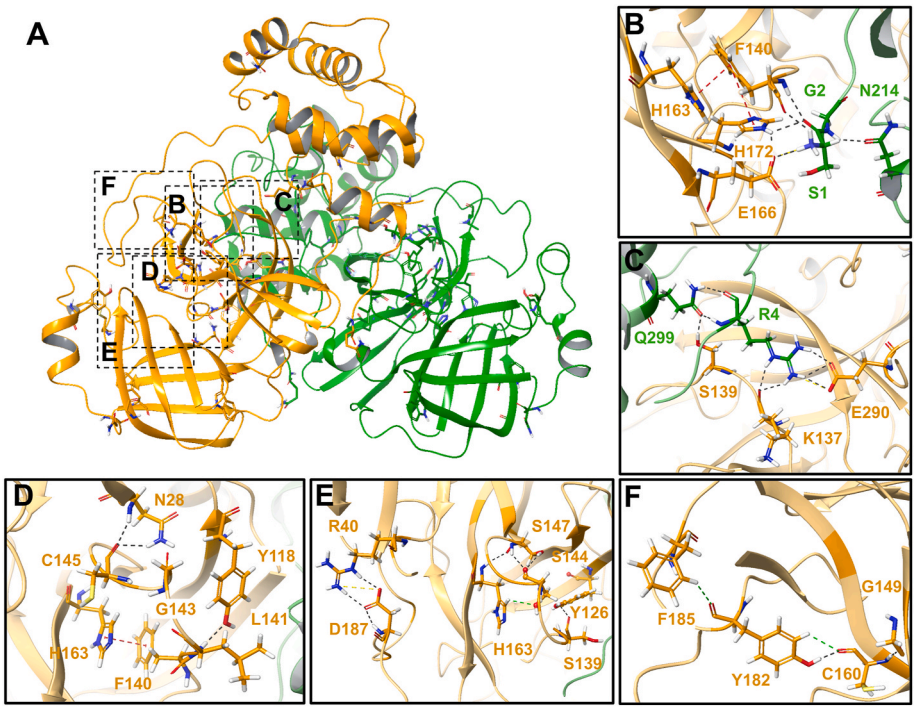

Fig. 3. Structural implications and non-covalent interactions of SARS-CoV-2 M-pro mutation coldspots. Main panel (A) shows the distribution of mutation coldspots in the M-pro dimeric structure. One protomer is green and the other is ocher. Accessory panels (B–F) show details of interactions between coldspots and/or other important residues. Interactions are depicted as discontinuous lines and colored according to the type of interaction: H-bond (dark gray), salt bridge (yellow), π-π stacking (red), aromatic H-bond (green). (B) Dimer interface interactions of the mutation coldspots Gly2, Phe140, His163 and His172, and the conserved residue Glu166. (C) Interaction network between Arg4 from one protomer and Glu290 and Lys137 from the other protomer. (D) Interactions of mutation coldspots Asn28 and Tyr118. (E) Interactions of mutation coldspots Ser139 and Ser147 and Ser144, near the substrate-binding site. (F) Interaction network between the mutation coldspot Tyr182 and neighboring residues.

crystal structures of SARS-CoV-2 M-pro, Glu166 also forms a hydrogen bond with His172 (Fig. 3B) [62]. Our definition of His172 as a mutation coldspot agrees with experimental results that show that M-pro is highly sensitive to changes in this residue [53,54] (Table 1). The substitution H172Y has been described to have a similar [63] or reduced [36] activity to that of the WT enzyme. But interestingly, this mutation causes the largest reduction in the inhibitory activity, with a 233-fold increase in the Ki value, of the M-pro inhibitor nirmatrelvir [64], and must be closely monitored among circulating SARS-CoV-2 viruses [36].

Residues near the substrate-binding site which are also involved in dimerization, such as Phe140 and Gly2, might be the link between dimerization and catalysis [65]. Thus, mutations of these residues collapse the substrate-binding site so that the enzyme is unable to perform catalysis. Correct positioning of Phe140 plays a pivotal role in the hydrophobic pocket formed with Tyr126 [66], identified here as a mutation coldspot (Table 1). The F140A mutation results in a dimeric M-pro conformation with a completely collapsed substrate-binding site [67] and a proteolytic activity of less than 1 % of the wild type [68]. The absence of a hydrogen bond interaction between Phe140 from one protomer and Ser1 from the other protomer produces a marginally active state of the main protease by destabilizing the oxyanion loop [69]. Gly2, a mutation coldspot (Table 1), is involved in an important interaction, linking the known allosteric coupling between domain III and the oxyanion loop [69]. Gly2 takes part in an interprotomer hydrogen bond with Asn214 (Fig. 3B). The absence of this interaction opens the dimer interface in the domain III of each protomer [69]. Note that Asn214 is not a mutation coldspot but only 25 missense mutations were found in more than 5,700,000 SARS-CoV-2 genomes analyzed, supporting the functional relevance of the Gly2-Asn214 interaction. Another residue from the N-finger that participates in dimerizarion is Phe3. The aromatic side chain of this residue is encapsulated in the hydrophobic pocket composed of side chains of Trp207, Ala210, Phe291 and Val296 [54,67]. Phe3 is not a mutational coldspot because the M-pro enzymatic activity is preserved when this residue is mutated to other hydrophobic residues [54]. Trp207 and Phe291 are mutation coldspots (Table 1), corroborating the structural role of these residues.

One well-known interaction of the dimer interface involves a salt bridge between the side chains of Arg4 and Glu290, each from a different protomer (Fig. 3C) [11,70]. This interaction is critical for SARS-CoV-2 M-pro function [71]. Glu290 is a mutation coldspot and mutations in this residue have a dramatic effects on the M-pro activity and dimerization [53,70–73] (Table 1). Truncation of residues 1–4 of the N-terminus of SARS-CoV M-pro resulted in the inability to dimerize and poor catalytic activity [74]. Interestingly, truncation of only the first three residues had only a limited effect on the protease. The M-pro truncated at these residues retained 76 % of the wild-type enzyme activity, reinforcing the idea that Arg4 is indeed an important residue within the N-terminal region [74]. However, Arg4 mutations are better tolerated than Glu290 mutations, and Arg4 is not a mutation coldspot. The R4A mutation causes a five-fold decrease in M-pro dimerization, but a moderate effect on the catalytic efficiency [72,73]. In addition, Arg4 also interacts with the backbone carbonyl oxygen of Lys137 of the oxyanion loop (Fig. 3C) [71]. The most common missense mutation of Arg4 (found 700 times) changes the Arg to a Lys, which may not affect the interactions of this residue with Glu290 and Lys137. In fact, the substitution to a lysine could induce tighter packing of the M-pro dimer, due to effect of the smaller residue, and therefore the R4K mutation increases the M-pro activity [53]. The destabilization of the Arg4/-Glu290 interaction has been suggested as a new strategy to design anti-SARS-CoV-2 compounds targeting the dimer interface [71]. However, recent studies have shown that disrupting only the Arg4/Glu290 salt bridge is insufficient to inhibit the formation of the SARS-CoV-2 M-pro dimer [75].

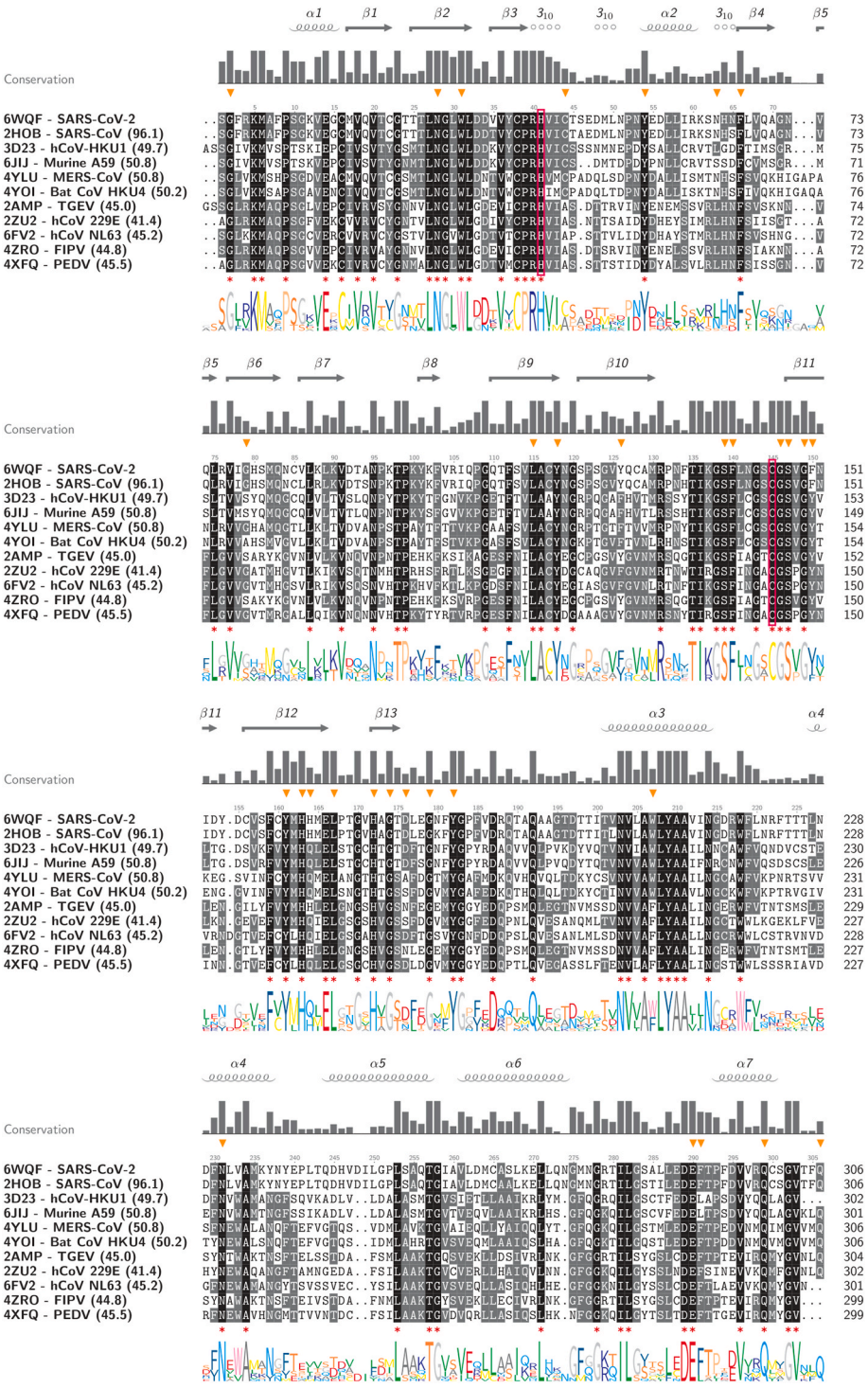

Fig. 4. Multiple sequence alignment of SARS-CoV-2 and other CoV M-pros. Each sequence of the multiple sequence alignment (MSA) is identified with the PDBid of a CoV main protease structure, the name of the virus and the identity percentage calculated by pairwise alignment between each sequence and the SARS-CoV-2 M-pro (in brackets). The secondary structure is displayed at the top of the MSA, along with the degree of conservation represented by a bar plot. Mutation coldspots are labelled using orange down-facing triangles. Totally conserved residues in all sequences are highlighted in black, whereas residues conserved in at least 50 % of sequences are highlighted in gray. A red asterisk also indicates totally conserved residues. Residues from the catalytic dyad are highlighted using a red rectangle. A sequence logo is displayed at the bottom and residues. Abbreviations: α, alpha helix; β, beta strand; 310, 310 helix; hCoV, human CoV; TGEV, Transmissible gastroenteritis virus; FIPV, Feline infectious peritonitis virus; PEDV, Porcine epidemic diarrhea virus.

3.2.2. Interaction networks of the substrate-binding site and its vicinity

Asn28, another mutation coldspot (Table 1), mediates key interactions within the M-pro structure, including two interactions with the backbone carbonyl of the catalytic residue Cys145 (Fig. 3D) [76]. In fact, mutation of Asn28 results in complete loss of enzymatic activity –probably due to its role in correct positioning of Cys145– and a dramatic increase (*i.e.,* roughly 20-fold) in the dissociation constant [66]. Specifically, the mutation N28A rotates the catalytic Cys145 side chain 180° to form a novel disulfide bond with Cys117 and affects the structure of the loop formed by residues 139–141 [66]. Asn28 is a good

example of a long-range interacting residue within M-pro. Mutation of this residue to Ala affects dimerization, even though this residue is located ~11 Å from any residue of the other protomer [66]. Kidera et al. [69], suggested that a hydrogen bond between Asn28 and the backbone carbonyl of Gly143 –which is not a mutation coldspot, but is a conserved residue (Table S3) extremely sensitive to mutations [53] – regulates the formation of the oxyanion hole in SARS-CoV-2 M-pro. In fact, both Gly143-Asn28 and Cys145-Asn28 hydrogen bonds are required in the correct arrangement of NH groups from residues 143 and 145 to correctly form the oxyanion hole [69]. The formation of the hydrogen bond between Asn28 and Gly143 is indeed a marker of the state of the protease: when the hydrogen bond is present the active state occurs, and it turns out to be collapsed when it disappears [69]. Kidera et al. [69] have also identified an hydrogen bond between the side chain of Tyr118 –a mutation coldspot (Table 1)– and the backbone amide of Leu141 (Fig. 3D) as a typical feature of active unbound conformations of the M-pro which can be barely found in the inactive state. Therefore, most mutations of Tyr118 affect the M-pro activity [53]. Bacha et al. [77] identified a cluster of three conserved serine residues in the main proteases of different CoVs: Ser139, Ser144 and Ser147. These residues are important in binding site shaping and orientation of the ligand. Mutation in any of these three residues has a significant impact on enzymatic activity [76]. In this analysis, we found that Ser139 and Ser147 are mutation coldspots (Table 1). Regarding Ser144, we have identified a total of three distinct missense mutations, namely S144A, S144L and S144E, each occurring at low frequencies (all of them 17 times). The backbone carbonyl of Ser139 interacts with the side chain of Tyr126 (Fig. 3E), another mutation coldspot (Table 1) [69]. In addition, the side chain of Ser139 forms an interprotomer hydrogen bond with the side chain of Gln299 (Fig. 3C), a mutation coldspot (Table 1) [17,70,71]. Mutations in Gln299 are sufficient to disrupt SARS-CoV M-pro dimerization [73] and increase the dissociation constant of the wild-type M-pro up to 4000-fold [78]. More recently, it was reported that only residual activity of the SARS-CoV-2 M-pro was observed when Gln299 was mutated to either Ala or Asn [71]. The S139A mutation in the M-pro of SARS-CoV disrupts the dimerization interface, weakening the dimerization constant by about 10-fold [76]. However, it does not interfere directly in the catalytic activity of the enzyme [76]. The backbone carbonyl of Ser144 interacts with the backbone amide of Ser147 (Fig. 3E) [76]. However, the S144A mutation in the SARS-CoV M-pro leads to a more structurally stable enzyme [76]. Perhaps for this reason, we have not observed Ser144 as a mutation coldspot. Concerning Ser147, it is a residue which is buried between domains I and II, just behind the binding pocket and it is located far away (~9 Å) from the dimer interface [76]. Its backbone carbonyl interacts with the backbones of Ser144 and His163 (Fig. 3E) [76]. The S147A mutation decreases about 150-fold the enzymatic activity of SARS-CoV M-pro [76]. However, mutations at Ser147 disrupt also dimerization [53], suggesting a long-cooperative interaction network between the dimer interface and the substrate-binding site [53,76].

His163 and His164 are substrate-binding site residues that are mutation coldspots (Table 1). In silico mutagenesis of His163 showed reduced affinity for ligands in molecular dynamics simulations, suggesting that it plays a prominent role in ligand binding [79]. His163 interacts via π-π stacking with Phe140 [54] (Fig. 3B and D) and forms a hydrogen bond with the side-chain hydroxyl of Tyr161 [60,63,67]. Phe140 and Tyr161 are mutation coldspot (Table 1) and both residues show low mutation tolerance [53,54,63]. Given its importance in ligand binding as well as its conservation, His163 is a noteworthy residue in the design of M-pro inhibitors [80]. The substitution H163A shows an inactive oxidized conformation of the M-pro, with the catalytic cysteine forming a disulfide bond with Cys117 [81], that is also seen in the crystal structure of the N28A mutant in the SARS-CoV M-pro [66]. His164 is a residue intolerant to mutations [54] that interacts with H_2O_{cat}, and it is conserved among CoVs (Fig. 4). The H164L mutation was found to have a mild effect on protein stability, but with a destabilizing effect [82]. The same in silico M-pro protein stability analysis showed that the L167S mutation results in a significant protein instability [82]. Here, we report that Leu167 is a mutation coldspot (Table 1). The L167F substitution decreases M-pro activity to around 20 % of that of the wild type [83,84], but also confers resistance to the M-pro inhibitor nirmatrelvir [58,83]. Gaining resistance at the expense of decreased enzymatic activity is a common mechanism among nirmatrelvir resistance mutations [58]. Although Leu167 is not directly involved in nirmatrelvir binding, the presence of a bulkier side chain, such as Phe, can cause a distortion of the binding pocket, resulting in a decrease of the van der Waals forces between the M-pro and its inhibitor [83]. Two other residues situated in the vicinity of the substrate-binding site are also mutation coldspots: Cys44 and Phe150 (Table 1). On one hand, Cys44 is part of a flexible loop composed of residues 44–52, which potentially regulates the access to the active site [85]. Mutations of the residues of this loop could notably change M-pro binding properties to inhibitors [85]. Cys44 may form a post-translational cross-link with Cys22 and Lys61, called SONOS bridge, that serves potentially as a redox switch to regulate the M-pro catalytic activity under oxidative stress conditions [86,87]. Cys22 is a conserved residue (Table S3) and the conservation of both cysteine residues (Cys44 and Cys22) may indicate their structural stability role under oxidative stress conditions. On the other hand, Phe150 is a critical residue involved in binding site shaping [81] and this residue exhibits low mutation tolerance (Table 1) [53]. Mutating this residue to larger or more polar ones causes a reduction in the M-pro activity [54]. Other residues that perform an indispensable function during the catalytic process in SARS-CoV M-pro are Arg40 and Tyr54 [88]. Tyr54 is defined here as a mutation coldspot and is one of the two M-pro residues for which no missense mutations have been observed (Table 1). Arg40 is defined as a conserved residue with five missense mutations, found eleven times in total (Table S3). Arg40 forms a salt bridge with Asp187 [53,54] (Fig. 3E) and exhibits low mutation tolerance [53]. Mutations Y54A and Y54C have been associated with resistance to the M-pro inhibitors nirmatrelvir and ensitrelvir [35,84]. The substitution of Tyr54 with a smaller residue, such as Cys, breaks a strong hydrogen bond with nirmatrelvir, frees up space in the M-pro binding pocket and may lead to a restructuring of the catalytic site, thereby indirectly affecting inhibitor binding [84].

3.2.3. Residues of the C-terminal domain

Tyr182, located in the M-pro C-terminal domain, is a residue defined here as a mutation coldspot and, along with Tyr54, is one of the two residues where no missense mutations have been observed (Table 1). Despite the lack of information about the role of this residue, its interaction with Phe185 (Fig. 3F), a conserved residue (Table S3), may contribute to shaping the binding site by correctly positioning the flexible loop Phe185-Thr201, which has been demonstrated to be crucial for SARS-CoV M-pro activity [89]. The side chain of Tyr182 also interacts with Cys160, which in turn interacts with Gly149 (Fig. 3F), defined here as a mutation coldspot (Table 1). Gly149 is one of the residues with low mutation tolerance in the M-pro protein and is located in a highly conserved region across other CoVs [53]. The strict conservation of Tyr182 may be attributed to its interactions with the residues mentioned earlier, which could be crucial for the protein's activity. The last residue of the M-pro, the Gln306 residue, is also a mutation coldspot, which has only undergone one missense mutation, found in only one genome (Table 1). This is expected due to the importance of this residue, which serves as a cleavage site recognized by M-pro itself to process the polyproteins pp1a and pp1ab into several functional proteins.

3.2.4. Mutation coldspots are widely conserved in different CoVs

Twenty-two out of 32 SARS-CoV-2 M-pro mutation coldspots defined in this manuscript are totally conserved in other CoVs, and 28 out of 32 are very conserved (Fig. 4). Note that mutation coldspots Asn63, Tyr126, His164 and Phe150 are different from the prevailing residue at

this position in the multiple sequence alignment, but are conserved in SARS-CoV and SARS-CoV-2 (Fig. 4). Interestingly, the coldspot Asn63 is only found in SARS-CoV, MERS-CoV and Bat CoV HKU4 M-pro (Fig. 4). Since this residue is located in the outer and more distal part of domain I in the M-pro structure, a clear function cannot be inferred. Most mutations of this residue do not affect the M-pro enzymatic activity (Table 1). Although this conservation may just be due to chance, a highly conserved residue after four years of evolution must have an important function, as is the case with most coldspots. This also applies to Phe150 (Tyr in all other CoVs apart from SARS-CoV and SARS-CoV-2). Phe150 exhibits a low mutation tolerance (Table 1), but the substitution of this residue to another aromatic residue, Tyr or Trp, even increases M-pro activity [53]. Of special interest is Leu286. I286L is one of the 12 substitutions found between the M-pro of SARS-CoV and SARS-CoV-2. This residue is highly variable in the aligned sequences of other CoVs (Fig. 4), suggesting a systematic mutation pattern at this position. Even though Leu286 is not a mutation coldspot, it is a conserved residue (*i.e.,* there are two different missense mutations annotated in this residue found only in six SARS-CoV-2 genomes) (Table S3). That might indicate that this residue was specifically and evolutionally selected to induce tighter packing of the M-pro dimer, and it is thereby well-conserved in SARS-CoV-2 M-pro. Among the conserved residues in almost all CoVs there is a conserved GSCGSxG motif (residues 143–149 of the reference SARS-CoV-2 M-pro sequence, see Fig. 4), which was identified as important for initiating catalysis in SARS-CoV and MERS-CoV [88]. Three residues of this motif (*i.e.*, Gly146, Ser147 and Gly149) are mutation coldspots (Table 1). Interestingly, mutation of Gly146 to Pro showed a significant decline in bioactivity in SARS-CoV M-pro, indicating the irreplaceability of this glycine residue in this motif [88]. Mutations in this residue reduce M-pro enzymatic activity and this residue is one of the key contributors to protein structure stability (Table 1).

4. Conclusion

The present study paves the way for further studies on SARS-CoV-2 genomic analyses. A total of 32 mutation coldspots were defined for the SARS-CoV-2 M-pro. It has been shown that some residues that are not apparently important, enable interactions that end up having a pivotal role. This is supported by the conservation analysis, which reveals that critical residues within the M-pro structure (*i.e.*, mutation coldspots) have a significant structural role rather than belonging strictly to the substrate-binding site. Most of the mutation coldspots defined mainly mediate interprotomer interactions or funneling interaction networks from the substrate-binding site towards the dimerization surface and *vice versa*. This new paradigm discussed here suggests that inhibition of the M-pro dimerization could be an interesting alternative to classical substrate-binding site inhibition approaches [75]. Goyal and Goyal [90] addressed this issue in a review article stating that inhibition of dimerization with peptide-based inhibitors could be a better therapeutic strategy than small molecules. Thus, it has been found that several compounds or peptides can bind to the dimer interface and inhibit SARS-CoV-2 M-pro [91–94] and a novel SARS-CoV-2 M-pro dimer-based screening system has been developed [95]. It is worth highlighting that residues, such as Glu290, which were stated as interesting targets for M-pro dimerization inhibition are among the herein defined mutation coldspots. Even though the bulk of work still focuses on the inhibition of the M-pro active site, we strongly believe that these approaches might be an interesting alternative. Thus, our work provides more insight into residues that could be targeted in these methodologies and a solid foundation for the design and development of new SARS-CoV-2 M-pro dimerization inhibitors. Moreover, given that most of coldspots are conserved in different CoVs, it might be an effective strategy for combating other viruses with a potential pan-CoV activity.

CRediT authorship contribution statement

Pol Garcia-Segura: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Ariadna Llop-Peiró:** Visualization. **Nil Novau-Ferré:** Data curation. **Júlia Mestres-Truyol:** Writing – original draft, Visualization. **Bryan Saldivar-Espinoza:** Software, Data curation. **Gerard Pujadas:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Santiago Garcia-Vallvé:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens (gisaid.org/EPI_SET_240913zy), and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We kindly acknowledge the English language service of our university for the edition and review of the present manuscript. This work was supported by the project PID2022-138327OB-I00 financed by financed by the Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI)/10.13039/501100011033/FEDER, UE. AL-P is recipient of the predoctoral grant 2022PMF-INV-14 from the INVESTIGO call that is financed by the Next Generation EU program (through the Recovery and Resilience Facility initiative), the Public Service of State Employment (SEPE) from the Spanish Government and Universitat Rovira i Virgili.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.109344.

References

[1] J.F.-W. Chan, K.-H. Kok, Z. Zhu, H. Chu, K.K.-W. To, S. Yuan, K.-Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerg Microbes Infect 9 (2020) 221–236, https://doi.org/10.1080/22221751.2020.1719902.

[2] J. Singh, P. Pandit, A.G. McArthur, A. Banerjee, K. Mossman, Evolutionary trajectory of SARS-CoV-2 and emerging variants, Virol. J. 18 (2021) 166, https://doi.org/10.1186/s12985-021-01633-w.

[3] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, P.E. Rollin, S.F. Dowell, A.-E. Ling, C.D. Humphrey, W.-J. Shieh, J. Guarner, C.D. Paddock, P. Rota, B. Fields, J. DeRisi, J.-Y. Yang, N. Cox, J.M. Hughes, J.W. LeDuc, W.J. Bellini, L.J. Anderson, SARS Working Group, A novel coronavirus associated with severe acute respiratory syndrome, N. Engl. J. Med. 348 (2003) 1953–1966, https://doi.org/10.1056/NEJMoa030781.

[4] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D.M.E. Osterhaus, R.A. M. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, N. Engl. J. Med. 367 (2012) 1814–1820, https://doi.org/10.1056/NEJMoa1211721.

[5] B. Hu, H. Guo, P. Zhou, Z.-L. Shi, Characteristics of SARS-CoV-2 and COVID-19, Nat. Rev. Microbiol. 19 (2021) 141–154, https://doi.org/10.1038/s41579-020-00459-7.

[6] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (2020) 270–273, https://doi.org/10.1038/s41586-020-2012-7.

[7] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E. C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation

and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574, https://doi.org/10.1016/S0140-6736(20)30251-8.
[8] H. Tan, Y. Hu, P. Jadhav, B. Tan, J. Wang, Progress and challenges in targeting the SARS-CoV-2 papain-like protease, J. Med. Chem. 65 (2022) 7561–7580, https://doi.org/10.1021/acs.jmedchem.2c00303.
[9] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.
[10] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293, https://doi.org/10.1038/s41586-020-2223-y.
[11] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Science (New York, N.Y.) 368 (2020) 409–412, https://doi.org/10.1126/science.abb3405.
[12] R. Cannalire, C. Cerchia, A.R. Beccari, F.S. Di Leva, V. Summa, Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: state of the art and future opportunities, J. Med. Chem. 65 (2022) 2716–2746, https://doi.org/10.1021/acs.jmedchem.0c01140.
[13] A. Gimeno, J. Mestres-Truyol, M.J. Ojeda-Montes, G. Macip, B. Saldivar-Espinoza, A. Cereto-Massagué, G. Pujadas, S. Garcia-Vallvé, Prediction of novel inhibitors of the main protease (M-pro) of SARS-CoV-2 through consensus docking and drug reposition, Int. J. Mol. Sci. 21 (2020) 3793, https://doi.org/10.3390/ijms21113793.
[14] G. Macip, P. Garcia-Segura, J. Mestres-Truyol, B. Saldivar-Espinoza, G. Pujadas, S. Garcia-Vallvé, A review of the current landscape of SARS-CoV-2 main protease inhibitors: have we hit the bullseye yet? Int. J. Mol. Sci. 23 (2021) 259, https://doi.org/10.3390/ijms23010259.
[15] G. Macip, P. Garcia-Segura, J. Mestres-Truyol, B. Saldivar-Espinoza, M.J. Ojeda-Montes, A. Gimeno, A. Cereto-Massagué, S. Garcia-Vallvé, G. Pujadas, Haste makes waste: a critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition, Med. Res. Rev. 42 (2022) 744–769, https://doi.org/10.1002/med.21862.
[16] D.W. Kneller, G. Phillips, H.M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky, Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography, Nat. Commun. 11 (2020) 3202, https://doi.org/10.1038/s41467-020-16954-7.
[17] D. Suárez, N. Díaz, SARS-CoV-2 main protease: a molecular dynamics study, J. Chem. Inf. Model. 60 (2020) 5815–5831, https://doi.org/10.1021/acs.jcim.0c00575.
[18] R. Wang, Y. Hozumi, Y.-H. Zheng, C. Yin, G.-W. Wei, Host immune response driving SARS-CoV-2 evolution, Viruses 12 (2020) 1095, https://doi.org/10.3390/v12101095.
[19] S.L. Kosakovsky Pond, D. Martin, Anti-COVID drug accelerates viral evolution, Nature 623 (2023) 486–487, https://doi.org/10.1038/d41586-023-03248-3.
[20] K.K. Irwin, N. Renzette, T.F. Kowalik, J.D. Jensen, Antiviral drug resistance as an adaptive process, Virus Evol 2 (2016) vew014, https://doi.org/10.1093/ve/vew014.
[21] B. Saldivar-Espinoza, P. Garcia-Segura, N. Novau-Ferré, G. Macip, R. Martínez, P. Puigbò, A. Cereto-Massagué, G. Pujadas, S. Garcia-Vallve, The mutational landscape of SARS-CoV-2, Int. J. Mol. Sci. 24 (2023) 9072, https://doi.org/10.3390/ijms24109072.
[22] S. Duffy, L.A. Shackelton, E.C. Holmes, Rates of evolutionary change in viruses: patterns and determinants, Nat. Rev. Genet. 9 (2008) 267–276, https://doi.org/10.1038/nrg2323.
[23] E. Mohammadi, F. Shafiee, K. Shahzamani, M.M. Ranjbar, A. Alibakhshi, S. Ahangarzadeh, L. Beikmohammadi, L. Shariati, S. Hooshmandi, B. Ataei, S. H. Javanmard, Novel and emerging mutations of SARS-CoV-2: biomedical implications, Biomed. Pharmacother. 139 (2021) 111599, https://doi.org/10.1016/j.biopha.2021.111599.
[24] L.D. Eckerle, M.M. Becker, R.A. Halpin, K. Li, E. Venter, X. Lu, S. Scherbakova, R. L. Graham, R.S. Baric, T.B. Stockwell, D.J. Spiro, M.R. Denison, Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing, PLoS Pathog. 6 (2010) e1000896, https://doi.org/10.1371/journal.ppat.1000896.
[25] R.S. Harris, J.P. Dudley, APOBECs and virus restriction, Virology (2015) 131–145, https://doi.org/10.1016/j.virol.2015.03.012, 479–480.
[26] E. Eisenberg, E.Y. Levanon, A-to-I RNA editing - immune protector and transcriptome diversifier, Nat. Rev. Genet. 19 (2018) 473–490, https://doi.org/10.1038/s41576-018-0006-1.
[27] M. Pachetti, B. Marini, F. Benedetti, F. Giudici, E. Mauro, P. Storici, C. Masciovecchio, S. Angeletti, M. Ciccozzi, R.C. Gallo, D. Zella, R. Ippodrino, Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant, J. Transl. Med. 18 (2020) 179, https://doi.org/10.1186/s12967-020-02344-6.
[28] C.L.D.C. Badua, K.A.T. Baldo, P.M.B. Medina, Genomic and proteomic mutation landscapes of SARS-CoV-2, J. Med. Virol. 93 (2021) 1702–1721, https://doi.org/10.1002/jmv.26548.
[29] X. Chen, S. Leyendecker, H. van den Bedem, SARS-CoV-2 main protease mutation analysis via a kinematic method, Proteins 91 (2023) 1496–1509, https://doi.org/10.1002/prot.26543.
[30] J.T. Lee, Q. Yang, A. Gribenko, B.S. Perrin, Y. Zhu, R. Cardin, P.A. Liberator, A. S. Anderson, L. Hao, Genetic surveillance of SARS-CoV-2 Mpro reveals high sequence and structural conservation prior to the introduction of protease inhibitor paxlovid, mBio 13 (2022) e0086922, https://doi.org/10.1128/mbio.00869-22.
[31] B. Saldivar-Espinoza, G. Macip, P. Garcia-Segura, J. Mestres-Truyol, P. Puigbò, A. Cereto-Massagué, G. Pujadas, S. Garcia-Vallve, Prediction of recurrent mutations in SARS-CoV-2 using artificial neural networks, Int. J. Mol. Sci. 23 (2022) 14683, https://doi.org/10.3390/ijms232314683.
[32] P. Colson, H. Chaudet, J. Delerce, P. Pontarotti, A. Levasseur, J. Fantini, B. La Scola, C. Devaux, D. Raoult, Role of SARS-CoV-2 mutations in the evolution of the COVID-19 pandemic, J. Infect. 88 (2024) 106150, https://doi.org/10.1016/j.jinf.2024.106150.
[33] T. Dey, S. Chatterjee, S. Manna, A. Nandy, S.C. Basak, Identification and computational analysis of mutations in SARS-CoV-2, Comput. Biol. Med. 129 (2021) 104166, https://doi.org/10.1016/j.compbiomed.2020.104166.
[34] M.D. Sacco, Y. Hu, M.V. Gongora, F. Meilleur, M.T. Kemp, X. Zhang, J. Wang, Y. Chen, The P132H mutation in the main protease of Omicron SARS-CoV-2 decreases thermal stability without compromising catalysis or small-molecule drug inhibition, Cell Res. 32 (2022) 498–500, https://doi.org/10.1038/s41422-022-00640-y.
[35] J.D. Ip, A. Wing-Ho Chu, W.-M. Chan, R. Cheuk-Ying Leung, S.M. Umer Abdullah, Y. Sun, K. Kai-Wang To, Global prevalence of SARS-CoV-2 3CL protease mutations associated with nirmatrelvir or ensitrelvir resistance, EBioMedicine 91 (2023) 104559, https://doi.org/10.1016/j.ebiom.2023.104559.
[36] Y. Hu, E.M. Lewandowski, H. Tan, X. Zhang, R.T. Morgan, X. Zhang, L.M.C. Jacobs, S.G. Butler, M.V. Gongora, J. Choy, X. Deng, Y. Chen, J. Wang, Naturally occurring mutations of SARS-CoV-2 main protease confer drug resistance to nirmatrelvir, ACS Cent. Sci. 9 (2023) 1658–1669, https://doi.org/10.1021/acscentsci.3c00538.
[37] Y. Zhou, K.A. Gammeltoft, L.A. Ryberg, L.V. Pham, H.D. Tjørnelund, A. Binderup, C.R. Duarte Hernandez, C. Fernandez-Antunez, A. Offersgaard, U. Fahnøe, G.H. J. Peters, S. Ramirez, J. Bukh, J.M. Gottwein, Nirmatrelvir-resistant SARS-CoV-2 variants with high fitness in an infectious cell culture system, Sci. Adv. 8 (2022), https://doi.org/10.1126/sciadv.add7197 eadd7197.
[38] J. Ou, E.M. Lewandowski, Y. Hu, A.A. Lipinski, A. Aljasser, M. Colon-Ascanio, R. T. Morgan, L.M.C. Jacobs, X. Zhang, M.J. Bikowitz, P.R. Langlais, H. Tan, J. Wang, Y. Chen, J.S. Choy, A yeast-based system to study SARS-CoV-2 Mpro structure and to identify nirmatrelvir resistant mutations, PLoS Pathog. 19 (2023) e1011592, https://doi.org/10.1371/journal.ppat.1011592.
[39] S. Iketani, H. Mohri, B. Culbertson, S.J. Hong, Y. Duan, M.I. Luck, M. K. Annavajhala, Y. Guo, Z. Sheng, A.-C. Uhlemann, S.P. Goff, Y. Sabo, H. Yang, A. Chavez, D.D. Ho, Multiple pathways for SARS-CoV-2 resistance to nirmatrelvir, Nature 613 (2023) 558–564, https://doi.org/10.1038/s41586-022-05514-2.
[40] S. Khare, C. Gurry, L. Freitas, M.B. Schultz, G. Bach, A. Diallo, N. Akite, J. Ho, R. T. Lee, W. Yeo, G.C. Curation Team, S. Maurer-Stroh, GISAID's role in pandemic response, China CDC Wkly 3 (2021) 1049–1051, https://doi.org/10.46234/ccdcw2021.255.
[41] J. Delgado, L.G. Radusky, D. Cianferoni, L. Serrano, FoldX 5.0: working with RNA, small molecules and a new graphical interface, Bioinformatics 35 (2019) 4168–4169, https://doi.org/10.1093/bioinformatics/btz184.
[42] R.P. Joosten, F. Long, G.N. Murshudov, A. Perrakis, The PDB_REDO server for macromolecular structure model optimization, IUCrJ 1 (2014) 213–220, https://doi.org/10.1107/S2052252514009324.
[43] U. Bodenhofer, E. Bonatesta, C. Horejš-Kainrath, S. Hochreiter, msa: an R package for multiple sequence alignment, Bioinformatics 31 (2015) 3997–3999, https://doi.org/10.1093/bioinformatics/btv494.
[44] M. Heinig, D. Frishman, STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins, Nucleic Acids Res. 32 (2004) W500–W502, https://doi.org/10.1093/nar/gkh429.
[45] L. Jaroszewski, M. Iyer, A. Alisoltani, M. Sedova, A. Godzik, The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins, PLoS Comput. Biol. 17 (2021) e1009147, https://doi.org/10.1371/journal.pcbi.1009147.
[46] B. Saldivar-Espinoza, G. Macip, G. Pujadas, S. Garcia-Vallve, Could nucleocapsid be a next-generation COVID-19 vaccine candidate? Int. J. Infect. Dis. 125 (2022) 231–232, https://doi.org/10.1016/j.ijid.2022.11.002.
[47] T.J. Cross, G.R. Takahashi, E.M. Diessner, M.G. Crosby, V. Farahmand, S. Zhuang, C.T. Butts, R.W. Martin, Sequence characterization and molecular modeling of clinically relevant variants of the SARS-CoV-2 main protease, Biochemistry 59 (2020) 3741–3756, https://doi.org/10.1021/acs.biochem.0c00462.
[48] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics 34 (2018) 4121–4123, https://doi.org/10.1093/bioinformatics/bty407.
[49] C. Chakraborty, A.R. Sharma, M. Bhattacharya, G. Agoramoorthy, S.-S. Lee, Evolution, mode of transmission, and mutational landscape of newly emerging SARS-CoV-2 variants, mBio 12 (2021) e0114021, https://doi.org/10.1128/mBio.01140-21.
[50] M.S.A. Parvez, M.M. Rahman, M.N. Morshed, D. Rahman, S. Anwar, M.J. Hosen, Genetic analysis of SARS-CoV-2 isolates collected from Bangladesh: insights into the origin, mutational spectrum and possible pathomechanism, Comput. Biol. Chem. 90 (2021) 107413, https://doi.org/10.1016/j.compbiolchem.2020.107413.
[51] S. Ullrich, K.B. Ekanayake, G. Otting, C. Nitsche, Main protease mutants of SARS-CoV-2 variants remain susceptible to nirmatrelvir, Bioorg Med Chem Lett 62 (2022) 128629, https://doi.org/10.1016/j.bmcl.2022.128629.

[52] A. Hegyi, A. Friebe, A.E. Gorbalenya, J. Ziebuhr, Mutational analysis of the active centre of coronavirus 3C-like proteases, J. Gen. Virol. 83 (2002) 581–593, https://doi.org/10.1099/0022-1317-83-3-581.
[53] J.M. Flynn, N. Samant, G. Schneider-Nachum, D.T. Barkan, N.K. Yilmaz, C. A. Schiffer, S.A. Moquin, D. Dovala, D.N.A. Bolon, Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms, Elife 11 (2022) e77433, https://doi.org/10.7554/eLife.77433.
[54] S. Iketani, S.J. Hong, J. Sheng, F. Bahari, B. Culbertson, F.F. Atanaki, A.K. Aditham, A.F. Kratz, M.I. Luck, R. Tian, S.P. Goff, H. Montazeri, Y. Sabo, D.D. Ho, A. Chavez, Functional map of SARS-CoV-2 3CL protease reveals tolerant and immutable sites, Cell Host Microbe 30 (2022) 1354–1362.e6, https://doi.org/10.1016/j.chom.2022.08.003.
[55] J.C. Ferreira, S. Fadl, A.J. Villanueva, W.M. Rabeh, Catalytic dyad residues His41 and Cys145 impact the catalytic activity and overall conformational fold of the main SARS-CoV-2 protease 3-chymotrypsin-like protease, Front. Chem. 9 (2021) 692168, https://doi.org/10.3389/fchem.2021.692168.
[56] D. Jacot, T. Pillonel, G. Greub, C. Bertelli, Assessment of SARS-CoV-2 genome sequencing: quality criteria and low-frequency variants, J. Clin. Microbiol. 59 (2021) e0094421, https://doi.org/10.1128/JCM.00944-21.
[57] G.D. Noske, Y. Song, R.S. Fernandes, R. Chalk, H. Elmassoudi, L. Koekemoer, C. D. Owen, T.J. El-Baba, C.V. Robinson, Covid Moonshot Consortium, G. Oliva, A. S. Godoy, An in-solution snapshot of SARS-COV-2 main protease maturation process and inhibition, Nat. Commun. 14 (2023) 1545, https://doi.org/10.1038/s41467-023-37035-5.
[58] Y. Duan, H. Zhou, X. Liu, S. Iketani, M. Lin, X. Zhang, Q. Bian, H. Wang, H. Sun, S. J. Hong, B. Culbertson, H. Mohri, M.I. Luck, Y. Zhu, X. Liu, Y. Lu, X. Yang, K. Yang, Y. Sabo, A. Chavez, S.P. Goff, Z. Rao, D.D. Ho, H. Yang, Molecular mechanisms of SARS-CoV-2 resistance to nirmatrelvir, Nature 622 (2023) 376–382, https://doi.org/10.1038/s41586-023-06609-0.
[59] N. Krishnamoorthy, K. Fakhro, Identification of mutation resistance coldspots for targeting the SARS-CoV2 main protease, IUBMB Life 73 (2021) 670–675, https://doi.org/10.1002/iub.2465.
[60] S. Chen, T. Hu, J. Zhang, J. Chen, K. Chen, J. Ding, H. Jiang, X. Shen, Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations, J. Biol. Chem. 283 (2008) 554–564, https://doi.org/10.1074/jbc.M705240200.
[61] S.-C. Cheng, G.-G. Chang, C.-Y. Chou, Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease, Biophys. J. 98 (2010) 1327–1336, https://doi.org/10.1016/j.bpj.2009.12.4272.
[62] N. Verma, J.A. Henderson, J. Shen, Proton-coupled conformational activation of SARS coronavirus main proteases and opportunity for designing small-molecule broad-spectrum targeted covalent inhibitors, J. Am. Chem. Soc. 142 (2020) 21883–21890, https://doi.org/10.1021/jacs.0c10770.
[63] K. Al Adem, J.C. Ferreira, S. Fadl, W.M. Rabeh, pH profiles of 3-chymotrypsin-like protease (3CLpro) from SARS-CoV-2 elucidate its catalytic mechanism and a histidine residue critical for activity, J. Biol. Chem. 299 (2023) 102790, https://doi.org/10.1016/j.jbc.2022.102790.
[64] J. Clayton, V.M. de Oliveira, M.F. Ibrahim, X. Sun, P. Mahinthichaichan, M. Shen, R. Hilgenfeld, J. Shen, Integrative approach to dissect the drug resistance mechanism of the H172Y mutation of SARS-CoV-2 main protease, J. Chem. Inf. Model. 63 (2023) 3521–3533, https://doi.org/10.1021/acs.jcim.3c00344.
[65] J. Shi, J. Sivaraman, J. Song, Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease, J. Virol. 82 (2008) 4620–4629, https://doi.org/10.1128/JVI.02680-07.
[66] J. Barrila, S.B. Gabelli, U. Bacha, L.M. Amzel, E. Freire, Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CL(pro), Biochemistry 49 (2010) 4308–4317, https://doi.org/10.1021/bi1002585.
[67] T. Hu, Y. Zhang, L. Li, K. Wang, S. Chen, J. Chen, J. Ding, H. Jiang, X. Shen, Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure, Virology 388 (2009) 324–334, https://doi.org/10.1016/j.virol.2009.03.034.
[68] S. Chen, J. Zhang, T. Hu, K. Chen, H. Jiang, X. Shen, Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis, J. Biochem. 143 (2008) 525–536, https://doi.org/10.1093/jb/mvm246.
[69] A. Kidera, K. Moritsugu, T. Ekimoto, M. Ikeguchi, Allosteric regulation of 3CL protease of SARS-CoV-2 and SARS-CoV observed in the crystal structure ensemble, J. Mol. Biol. 433 (2021) 167324, https://doi.org/10.1016/j.jmb.2021.167324.
[70] A. Aniana, N.T. Nashed, R. Ghirlando, L. Coates, D.W. Kneller, A. Kovalevsky, J. M. Louis, Insights into the mechanism of SARS-CoV-2 main protease autocatalytic maturation from model precursors, Commun. Biol. 6 (2023) 1159, https://doi.org/10.1038/s42003-023-05469-8.
[71] J.C. Ferreira, S. Fadl, W.M. Rabeh, Key dimer interface residues impact the catalytic activity of 3CLpro, the main protease of SARS-CoV-2, J. Biol. Chem. 298 (2022) 102023, https://doi.org/10.1016/j.jbc.2022.102023.
[72] C.-Y. Chou, H.-C. Chang, W.-C. Hsu, T.-Z. Lin, C.-H. Lin, G.-G. Chang, Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease, Biochemistry 43 (2004) 14958–14970, https://doi.org/10.1021/bi0490237.
[73] J. Shi, J. Song, The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain, FEBS J. 273 (2006) 1035–1045, https://doi.org/10.1111/j.1742-4658.2006.05130.x.
[74] W.-C. Hsu, H.-C. Chang, C.-Y. Chou, P.-J. Tsai, P.-I. Lin, G.-G. Chang, Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease, J. Biol. Chem. 280 (2005) 22741–22748, https://doi.org/10.1074/jbc.M502556200.
[75] K. Lis, J. Plewka, F. Menezes, E. Bielecka, Y. Chykunova, K. Pustelny, S. Niebling, A.S. Garcia, M. Garcia-Alai, G.M. Popowicz, A. Czarna, T. Kantyka, K. Pyrc, SARS-CoV-2 Mpro oligomerization as a potential target for therapy, Int. J. Biol. Macromol. 267 (2024) 131392, https://doi.org/10.1016/j.ijbiomac.2024.131392.
[76] J. Barrila, U. Bacha, E. Freire, Long-range cooperative interactions modulate dimerization in SARS 3CLpro, Biochemistry 45 (2006) 14908–14916, https://doi.org/10.1021/bi0616302.
[77] U. Bacha, J. Barrila, A. Velazquez-Campoy, S.A. Leavitt, E. Freire, Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro, Biochemistry 43 (2004) 4906–4912, https://doi.org/10.1021/bi0361766.
[78] P.-Y. Lin, C.-Y. Chou, H.-C. Chang, W.-C. Hsu, G.-G. Chang, Correlation between dissociation and catalysis of SARS-CoV main protease, Arch. Biochem. Biophys. 472 (2008) 34–42, https://doi.org/10.1016/j.abb.2008.01.023.
[79] Y.L. Weng, S.R. Naik, N. Dingelstad, M.R. Lugo, S. Kalyaanamoorthy, A. Ganesan, Molecular dynamics and in silico mutagenesis on the reversible inhibitor-bound SARS-CoV-2 main protease complexes reveal the role of lateral pocket in enhancing the ligand affinity, Sci. Rep. 11 (2021) 7429, https://doi.org/10.1038/s41598-021-86471-0.
[80] D.D. Nguyen, K. Gao, J. Chen, R. Wang, G.-W. Wei, Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning, Chem. Sci. 11 (2020) 12036–12046, https://doi.org/10.1039/d0sc04641h.
[81] N. Tran, S. Dasari, S.A.E. Barwell, M.J. McLeod, S. Kalyaanamoorthy, T. Holyoak, A. Ganesan, The H163A mutation unravels an oxidized conformation of the SARS-CoV-2 main protease, Nat. Commun. 14 (2023) 5625, https://doi.org/10.1038/s41467-023-40023-4.
[82] E. Akbulut, Investigation of changes in protein stability and substrate affinity of 3CL-protease of SARS-CoV-2 caused by mutations, Genet. Mol. Biol. 45 (2022) e20210404, https://doi.org/10.1590/1678-4685-GMB-2021-0404.
[83] D. Jochmans, C. Liu, K. Donckers, A. Stoycheva, S. Boland, S.K. Stevens, C. De Vita, B. Vanmechelen, P. Maes, B. Trüeb, N. Ebert, V. Thiel, S. De Jonghe, L. Vangeel, D. Bardiot, A. Jekle, L.M. Blatt, L. Beigelman, J.A. Symons, P. Raboisson, P. Chaltin, A. Marchand, J. Neyts, J. Deval, K. Vandyck, The substitutions L50F, E166A, and L167F in SARS-CoV-2 3CLpro are selected by a protease inhibitor in vitro and confer resistance to nirmatrelvir, mBio 14 (2023) e0281522, https://doi.org/10.1128/mbio.02815-22.
[84] E. Heilmann, F. Costacurta, S.A. Moghadasi, C. Ye, M. Pavan, D. Bassani, A. Volland, C. Ascher, A.K.H. Weiss, D. Bante, R.S. Harris, S. Moro, B. Rupp, L. Martinez-Sobrido, D. von Laer, SARS-CoV-2 3CLpro mutations selected in a VSV-based system confer resistance to nirmatrelvir, ensitrelvir, and GC376, Sci. Transl. Med. 15 (2023), https://doi.org/10.1126/scitranslmed.abq7360 eabq7360.
[85] M. Bzówka, K. Mitusińska, A. Raczyńska, A. Samol, J.A. Tuszyński, A. Góra, Structural and evolutionary analysis indicate that the SARS-CoV-2 Mpro is a challenging target for small-molecule inhibitor design, Int. J. Mol. Sci. 21 (2020) 3099, https://doi.org/10.3390/ijms21093099.
[86] K.S. Yang, L.R. Blankenship, S.-T.A. Kuo, Y.J. Sheng, P. Li, C.A. Fierke, D. H. Russell, X. Yan, S. Xu, W.R. Liu, A novel Y-shaped, S-O-N-O-S-bridged cross-link between three residues C22, C44, and K61 is frequently observed in the SARS-CoV-2 main protease, ACS Chem. Biol. 18 (2023) 449–455, https://doi.org/10.1021/acschembio.2c00695.
[87] L.-M. Funk, G. Poschmann, F. Rabe von Pappenheim, A. Chari, K.M. Stegmann, A. Dickmanns, M. Wensien, N. Eulig, E. Paknia, G. Heyne, E. Penka, A.R. Pearson, C. Berndt, T. Fritz, S. Bazzi, J. Uranga, R.A. Mata, M. Dobbelstein, R. Hilgenfeld, U. Curth, K. Tittmann, Multiple redox switches of the SARS-CoV-2 main protease in vitro provide opportunities for drug design, Nat. Commun. 15 (2024) 411, https://doi.org/10.1038/s41467-023-44621-0.
[88] H. Wang, S. He, W. Deng, Y. Zhang, G. Li, J. Sun, W. Zhao, Y. Guo, Z. Yin, D. Li, L. Shang, Comprehensive insights into the catalytic mechanism of Middle East respiratory syndrome 3C-like protease and severe acute respiratory syndrome 3C-like protease, ACS Catal. 10 (2020) 5871–5890, https://doi.org/10.1021/acscatal.0c00110.
[89] M.-Y. Tsai, W.-H. Chang, J.-Y. Liang, L.-L. Lin, G.-G. Chang, H.-P. Chang, Essential covalent linkage between the chymotrypsin-like domain and the extra domain of the SARS-CoV main protease, J. Biochem. 148 (2010) 349–358, https://doi.org/10.1093/jb/mvq071.
[90] B. Goyal, D. Goyal, Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy, ACS Comb. Sci. 22 (2020) 297–305, https://doi.org/10.1021/acscombsci.0c00058.
[91] A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C.D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T.D. Downes, L. Dunnett, M. Fairhead, J.D. Firth, S.P. Jones, A. Keeley, G.M. Keserü, H.F. Klein, M.P. Martin, M.E.M. Noble, P. O'Brien, A. Powell, R.N. Reddi, R. Skyner, M. Snee, M.J. Waring, C. Wild, N. London, F. von Delft, M.A. Walsh, Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease, Nat. Commun. 11 (2020) 5047, https://doi.org/10.1038/s41467-020-18709-w.
[92] A.A. Panagiotopoulos, I. Karakasiliotis, D.-M. Kotzampasi, M. Dimitriou, G. Sourvinos, M. Kampa, S. Pirintsos, E. Castanas, V. Daskalakis, Natural polyphenols inhibit the dimerization of the SARS-CoV-2 main protease: the case of fortunellin and its structural analogs, Molecules 26 (2021) 6068, https://doi.org/10.3390/molecules26196068.

[93] S. Song, Y. Kim, K. Kwak, H. Lee, H. Park, Y.B. Kim, H.J. Lee, L.-W. Kang, The N-terminal peptide of the main protease of SARS-CoV-2, targeting dimer interface, inhibits its proteolytic activity, BMB Rep (2023) 6032.

[94] X. Tao, L. Zhang, L. Du, R. Liao, H. Cai, K. Lu, Z. Zhao, Y. Xie, P.-H. Wang, J.-A. Pan, Y. Zhang, G. Li, J. Dai, Z.-W. Mao, W. Xia, Allosteric inhibition of SARS-CoV-2 3CL protease by colloidal bismuth subcitrate, Chem. Sci. 12 (2021) 14098–14102, https://doi.org/10.1039/d1sc03526f.

[95] E.A. Giri-Rachman, V.V. Effendy, M.H.S. Azmi, N. Yamahoki, R. Stephanie, D. F. Agustiyanti, P.H. Wisnuwardhani, M. Angelina, Y. Rubiyana, R. Aditama, R. A. Ningrum, A. Wardiana, A. Fibriani, The SARS-CoV-2 Mpro dimer-based screening system: a synthetic biology tool for identifying compounds with dimerization inhibitory potential, ACS Synth. Biol. 13 (2024) 509–520, https://doi.org/10.1021/acssynbio.3c00446.